Abstract

Giant viruses belong to the phylum Nucleocytoviricota, a diverse group that comprises both isolated genomes (IGs) and giant virus metagenome-assembled genomes (GVMAGs). The diversity and gene content of giant viruses were analyzed using pangenome and gene-sharing network analyses. A large number of GVMAGs were recovered from marine environments through binning-based assembly methods, with IM_01 (Mesomimiviridae) representing the most abundant group. In contrast, recovery from soil environments was limited due to fragmented data. No environment-based patterns or clusters were observed in the phylogenetic tree. Functional annotation of the predicted genes revealed that most are involved in host metabolism and viral replication, with some unique functions, such as photosynthesis and multidrug resistance, also being identified. Pangenome analysis revealed varying levels of diversity across families, with some forming monophyletic groups and others exhibiting broader variation. Many GVMAGs lacking isolate genome representatives did not contain core genes, highlighting the importance of IGs in genome completeness and the need for taxonomic refinement. These patterns were often linked to host specificity. Gene interaction networks revealed distinct clustering across orders, further supporting the lineage-specific evolution of giant viruses.

Keywords*: giant virus, Giant Virus Metagenome-assembled genomes (GVMAGs), pangenome, Gene-sharing network, Viral Functional Annotation*